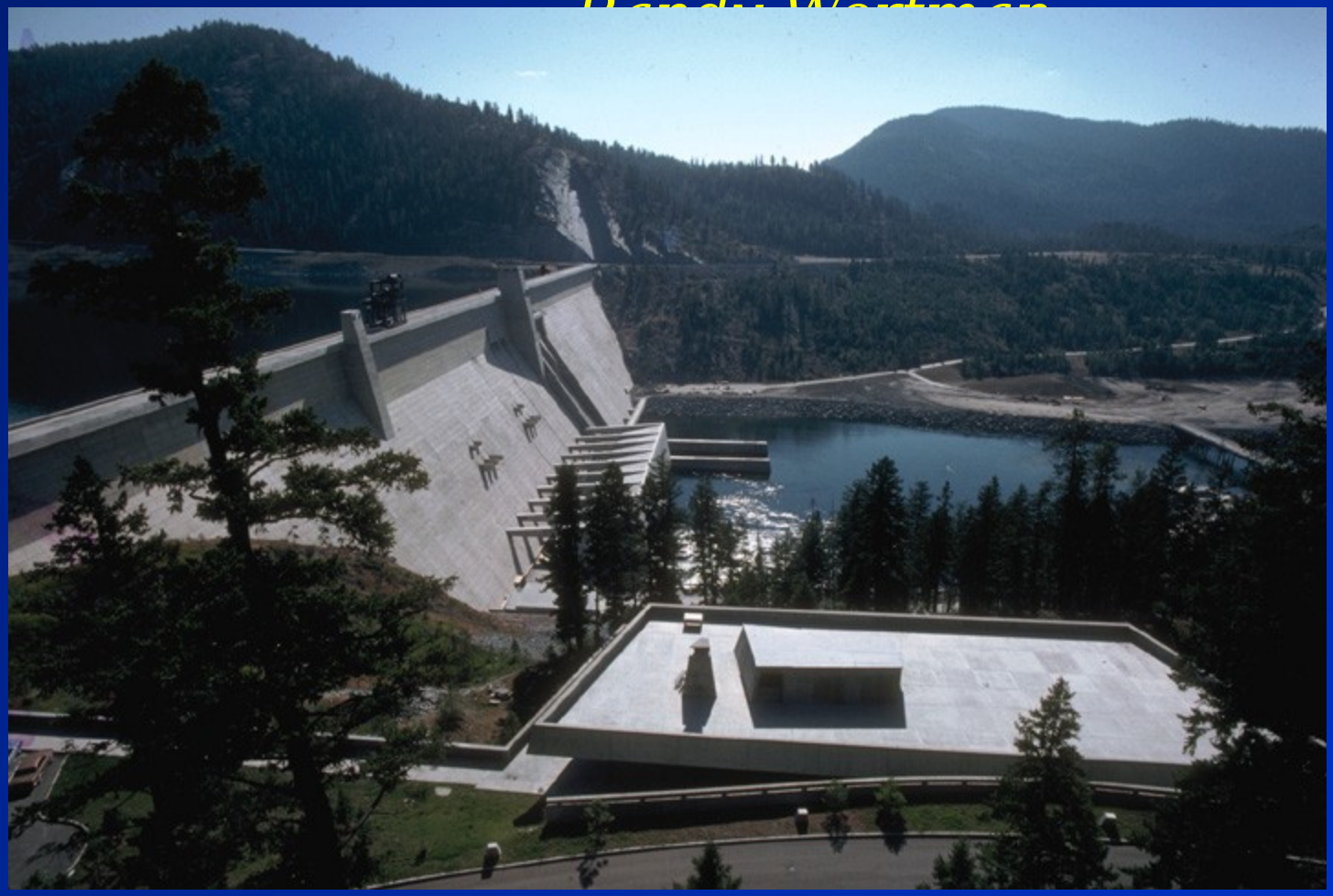
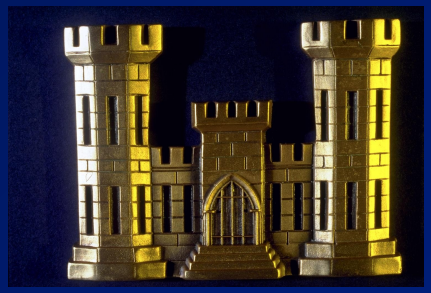
Update of Libby Statistical Forecast Procedure to Improve Early Season Forecasts - Does SOI Help?

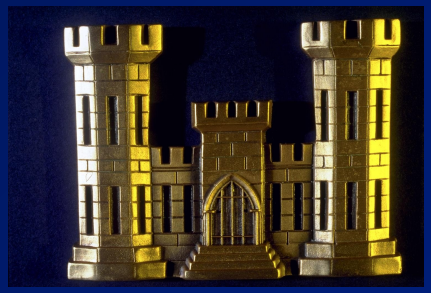
Randy Wortman





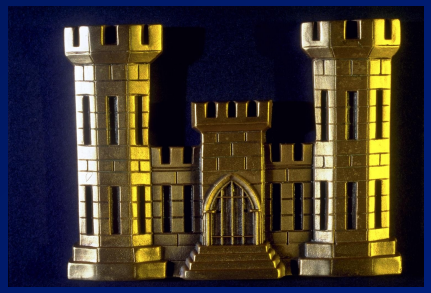
Current Libby Forecast Model

- Standard Linear Regression Equations
- Split Basin (2 Basins => 2 Regression Equations)
  - Libby Forecast = Ft. Steele Regr + Libby Local Regr
- Each equation utilizes four “surrogate” variables
  - Fall Runoff (Sum Oct-Dec subbasin runoff)
  - Winter Precipitation (Sum Oct-Mar precip at 4~5 sites)
  - April 1 SWE (Sum of SWE at 7 sites[Obs+Norm subsequent])
  - Spring Precip (Weighted Sum of Apr-May precip at 4~5 sites [Obs + Norm Sub])



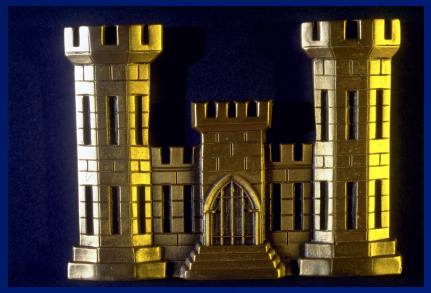
Current Model Advantages

- Traditional, familiar approach
- Somewhat conceptual
- Ease of use
- Positive correlation between recent precipitation and direction of forecast change
- “Good” month-to-month forecast consistency
- “Good” forecast accuracy



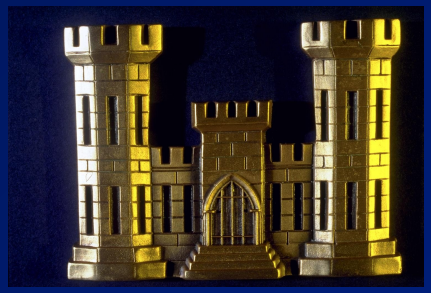
Current Model Concerns

- Variable selection technique: index variable components and weighting factors was highly subjective
- Intercorrelation of variables within each equation.
- Summation of two intercorrelated linear models
- Forecasting variables different than calibration variables (Use of “Normal Subsequent” variables will produce sub-optimal forecast)
- Model statistics (parsimony;degrees of freedom; Standard Error of Forecast - all difficult to compute or interpret)



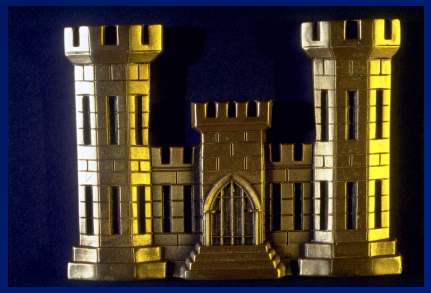
Current Model Concerns

- Model selection / fitting techniques suspect (minimizing the SE does not necessarily produce a good forecast model)
- Poor 1 January forecast
- No forecasts available prior to 1 January
- Other basin forecasts are showing improvement through use of an SOI variable



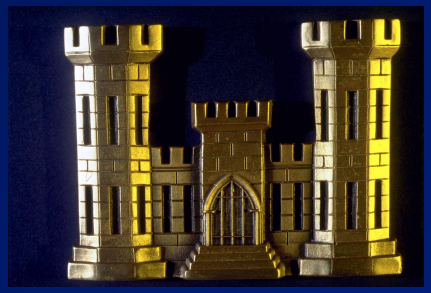
Current Investigation

- Principal Components Regression
  - Optimal variable weighting
  - Optimal variable selection
  - Uncorrelated variables
  - Parsimonious
  - Statistically sound



Current Investigation

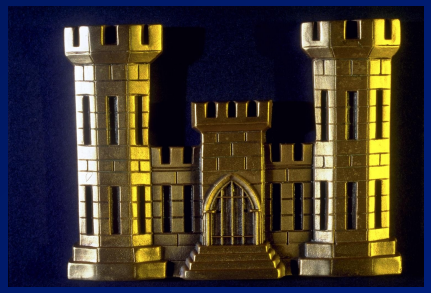
- Model Evaluation & Selection
  - Biased models result when variable selection and parameter estimation use the same dataset.
  - Cross-Validation Standard Error statistic
    - Provides a statistically sound model selection criteria
    - Intuitively mimics the “real world” - every forecast is made using a model where the parameters have been fit without using the data of the current observation.



Current Investigation - Data

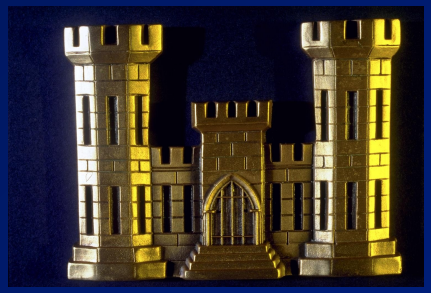
- SOI dataset: 1951-2000 available
  - How should monthly SOI values be handled?
    - Use individual monthly values or aggregate values

- Snow dataset:
  - No snow data prior to 1 January
  - 1 station with 40 years of 1 Jan SWE record
  - 5 stations with 21-30 years of 1 Jan SWE
  - 5 stations with 15-20 years of 1 Jan SWE



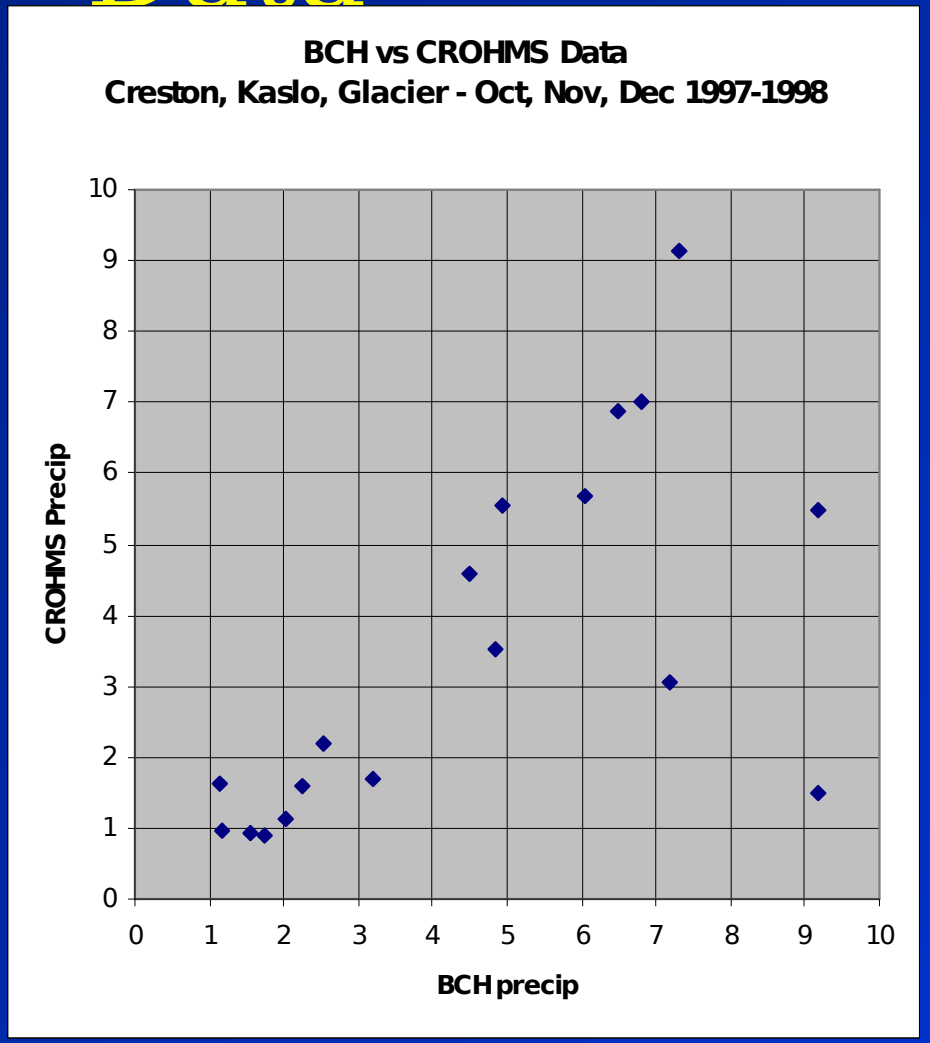
Current Investigation - Data

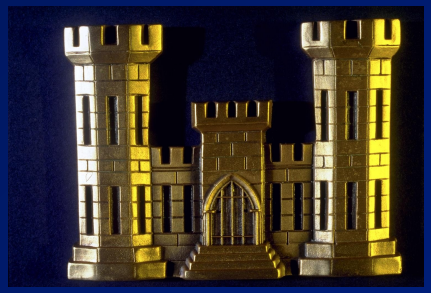
- Precipitation Dataset (Oct, Nov, Dec)
  - 5 stations with 1948-1999 data
    - Creston, Kaslo, Fernie, Wasa, Glacier RP
  - Brisco & Banff: 1948-1994 data
    - recent 4~5 years missing
  - Cranbrook: 1967-1999 data
    - No data prior to 1966



Current Investigation - Data

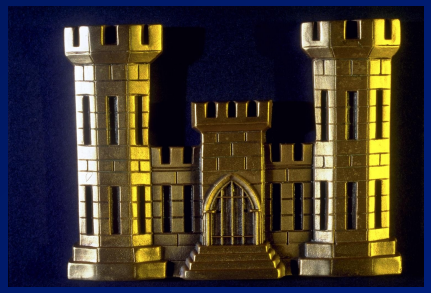
- Precipitation data received by CROHMS is frequently incomplete
- CROHMS monthly values vary greatly from Envir. Canada
- => CROHMS precip data is virtually worthless





Current Investigation-Modeling

		Forecast Period	N / PC	SOI Used?	1 January Forecast Std Error	CVSE
Split Basin Regression	1986 Update: 1948-85	Apr-Aug	38 / 0	No	1174	
NRCS	2000 Investigation	Apr-Jul	24 / 1	Yes	639	695
Corps	preliminary	Apr-Aug	43 / 3	No	918	1000
Corps	preliminary	Apr-Aug	43 / 2	Yes	899	936
Design		Apr-Aug	52 years	Yes	?	?



Next Steps

- Which SOI variables ?
- Fill out Brisco and Banff Precip (1995-99)
- Which snow data ?
  - Use only Nelson (1960-1999)
  - Include Marble Canyon, Fernie East, Moyie Mountain (mid 1970's - 1999)
  - Include US stations (Bonners Ferry)
- 1 December Forecast Model
- 1 November Forecast Model ??